UNCLASSIFIED MASTER COPY - FOR REPRODUCTION PURPOSES 2

SECURITY CLASSIFICATION OF THIS PAGE

AD-A204 773

DTIC ELECTE FEB 1 7 1989 S D D Cg

DTIC FILE COPY

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARO 23348.19-EL
6a. NAME OF PERFORMING ORGANIZATION Yale University / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code) 1504A Yale Station New Haven, CT 06520	7b. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U. S. Army Research Office / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER DAAL03-86-K-0158
8c. ADDRESS (City, State, and ZIP Code) P. O. Box 12211 Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification) Fast Design of Good Systolic Systems

12. PERSONAL AUTHOR(S) Jean-Marc Delosme, Ilse Ipsen

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 1 Sept 86 TO 31 Oct 88	89,01,12	

16. SUPPLEMENTARY NOTATION The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	VLSI technology, parallel architectures, systolic arrays, systolic hardware design tools, recurrence equations, parallel algorithms, signal processing

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Advanced high resolution methods for real-time signal processing require fast multi-processor architectures, realized as Very Large Scale Integrated systems, for their implementation. A class of parallel architectures, 'systolic arrays', fulfills the demands of signal processing and at the same time conforms to the limitations of VLSI technology.

The design of efficient systems of systolic arrays is an iterative process in which the information gathered from mapping algorithms onto hardware is influential in the development of the algorithms. Presently, the absence of mapping tools makes it extremely difficult to find good systolic implementations for many important problems. Often, one cannot do better than implementing structurally simple but numerically inferior algorithms, which is undesirable for most signal processing tasks.

A methodology is proposed that automates the mapping of recurrence equations to processor arrays. Two aspects distinguish our methodology from extant work: (1) complex coupled systems of recurrence equations can be systematically treated and (2) the resulting —

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED

19. systolic systems are optimal.

The proposed methodology takes as input recurrence equations describing the algorithm, along with certain desirable hardware features. A sophisticated optimization procedure is then applied to generate the description of optimal arrays that implement the algorithm. If there is no satisfactory implementation, the algorithm will have to be revised, in order to improve pipelinability, without sacrificing numerical stability.

Several classes of applications will significantly benefit from this approach: fast analysis of parallel algorithms, efficient partitioning of algorithms, and determination of optimal architectures for the implementation of multiple algorithms.

ARO 23348.19-EL

Fast Design of Good Systolic Systems

FINAL REPORT

J.-M. Delosme, I. Ipsen

January 12, 1989

U. S. ARMY RESEARCH OFFICE

DAAL03-86-K-0158

YALE UNIVERSITY

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED

A-1

A. Statement of Problem Studied

Systems for such applications as radar and sonar detection and identification, robot vision, medical imaging and geophysical exploration have to extract features from or make decisions based on large amounts of data. In order to minimize the computation time or to keep storage low, such systems often process data in *real time* and therefore consist of *many* processors. To be sufficiently fast, the processors are special purpose arithmetic computers, presently realized as VLSI chips. Because of their impact on the whole system, some limitations of the VLSI technology must be taken into account at the outset of the design process. The most important are the high design cost for developing novel processor architectures and the limited input/output capabilities of chips due to small pin counts. These constraints combined with the desire for efficient pipelining have led to the concept of 'systolic arrays'.

Systolic arrays consist of a few types of processors that perform elementary operations (such as additions or multiplications) selected by a simple control unit. The pattern of interconnections among processors is regular and each processor is connected to only a few other processors. Since the types of operations performed are essentially data independent, the processors can be made to operate synchronously without affecting the global performance; this simplifies the control of the processors and reduces the area needed to implement them in silicon.

The process of designing such systems can be decomposed into three steps :

- Development of signal processing techniques.
- Development of fast, stable and pipelinable algorithms that perform the optimization.
- Scheduling the algorithms onto minimal cost systolic architectures that deliver the desired throughput and comply with the current technological constraints. The best architecture may be a combination of several systolic arrays with different processors or interconnection patterns.

Clearly the three design steps are not completely independent and the best designs will be obtained through an iterative procedure; if there is no satisfactory implementation of an algorithm it will have to be revised. To make such iterations practically feasible the time consuming scheduling process must be automated and efficient schedules must be determined at every iteration. The class of algorithms to be implemented on systolic arrays is large, yet most of them can be expressed as systems of recurrence equations. The goal of the scheduling method is to construct a realistic systolic array model and to find optimal schedules (in this model) for *any* system of recurrence equations. By applying the method, while still under development, to the design of parallel implementations for algorithms, we can in turn refine the model of systolic algorithms and the objective functions that characterize good systolic systems.

B. Summary of the Most Important Results

Referring to the issues in Section 5 of our proposal, we can sum up our contributions supported under this ARO contract as follows:

1. **Preprocessing of Recurrence Equations.** We exploit the structure of an algorithm by decomposing its dependence graph into *strongly connected components*. By introducing the concept of *dependence mapping*, we generalize the established concept of *dependence vector* and can so represent many more signal processing algorithms.
2. **Determination of the Class of Local Transformations.** We have developed notions and procedures (e.g. time cones and computability analysis) to determine whether indeed an algorithm is well-defined and to provide a compact description of all systolic schedules. By applying

a combination of transformations from a restricted class (affine, folding and propagation transformations) we can transform the class of affine recurrences to the simpler class of uniform recurrences (with conditionals), which is directly amenable to systolic implementation.

The introduction of dependence mappings is a crucial first step in developing a systematic and efficient method for transforming broadcast dependences to propagation dependences that require a minimal communication bandwidth.

3. **Solution of the Global Optimization Problem and Determination of Local Transformations.** The time cone and the removal of broadcast dependences make it finally possible to solve the difficult problem of minimizing the computation time, more precisely: the product of systolic cycle time and number of cycles. Tools from integer linear programming theory, which we explored in the context of time optimization, turn out to provide tight bounds on the search space for optimal processor allocation schemes; hitherto no bounds were available and good solutions could be missed.

   The problem of finding a processor allocation that balances the load on each processor and minimizes the amount of interprocessor communication is highly combinatorial, especially for multistep algorithms. We have developed procedures for generating many possible processor allocation schemes; the problem left is to explore these possibilities and to select an (almost) optimum one: use of the simulated annealing method results in very high quality solutions. A long term investigation of simulated annealing resulted in a novel annealing schedule, significantly faster than any other in the literature. Using this schedule optimal allocations can now be obtained in a reasonable amount of time.

   In order to assure that the systolic array satisfies the required performance specifications (e.g. matching the specified data rate for real-time applications), we have explored 'partitioning strategies' for grouping computations to be mapped to the same processor. These were applied to schedule several algorithms onto the CMU WARP.

4. **Algorithm Design.** As our motivation is the development of parallel algorithms and efficient implementations, we have used the above methods to analyze the structure of algortihms suitable for systolic implementation. Conversely, the study of difficult algorithms has helped us to advance our work on scheduling and processor allocation. For example, we have investigated the parallel implemntation of singular value and symmetric eigenvalue problems, Toeplitz system solution, computation of partial correlations, and mesh refinement for partial differential equations.

   Further details can be found in the progress reports.

C. List of All Publications and Technical Reports

- Delosme, J.-M., *Systolic Models and Systolic Geometry*, Proc. Int. Symp. on Circuits and Systems ISCAS-88, Espoo, Finland, pp 2509-2512, June 1988.
- Delosme, J.-M., *A Processor for Two-Dimensional Symmetric Eigenvalue and Singular Value Arrays*, Proc. 21st Asilomar Conf. on Circuits, Systems and Computers, pp 217-221, November 1987.
- Delosme, J.-M., *A Processor for Two-Dimensional Symmetric Eigenvalue and Singular Value Arrays*, Research Report 540, Department of Computer Science, Yale University, May 1987.

  Accepted for publication in *IEEE Trans. on Computers*.
- Delosme, J.-M., Eisenstat, S.C. and Massé, J.R., *Toeplitz Solvers and Vector Processing*, Proc. 11th GRETSI Symposium on Signal and Image Processing, Nice, France, pp 665-668, June 1987.

- Delosme J.-M. and Ipsen, I.C.F., *Computing Partial Correlations from the Data Matrix*, Research Report 541, Department of Computer Science, Yale University, June 1987.
- Delosme, J.-M. and Ipsen, I.C.F., *From Bareiss' Algorithm to the Stable Computation of Partial Correlations*, Research Report 624, Department of Computer Science, Yale University, May 1988.

  Accepted for publication in *Journal of Computational and Applied Mathematics*.
- Delosme, J.-M. and Ipsen, I.C.F., *SAGA and CONDENSE: A Two-Phase Approach for the Implementation of Recurrence Equations on Multiprocessor Architectures*, Proc. 21st Hawai Int. Conf. on System Sciences, Vol. I, pp 126-130, January 1988.
- Delosme, J.-M., Ipsen, I.C.F., and Paige, C.C., *The Cholesky Factorization, Schur Complements, Correlation Coefficients, Angles between Vectors, and the QR Factorization*, Research Report 607, Department of Computer Science, Yale University, February 1988.
- Gropp, W.D. and Ipsen, I.C.F., *A Gray Code Scheme for Local Uniform Mesh Refinement on Hypercubes*, Proc. 3rd SIAM Conf. on Parallel Processing for Scientific Computing, SIAM, 1988.
- Gropp, W.D. and Ipsen, I.C.F., *Recursive Mesh Refinement on Hypercubes*, Research Report 616, Department of Computer Science, Yale University, March 1988.

  Submitted to *BIT*.
- Ipsen, I.C.F., *Systolic Algorithms for the Parallel Solution of Dense Symmetric Positive-Definite Toeplitz Systems*, Research Report 539, Department of Computer Science, Yale University, May 1987.

  Published in *Numerical Algorithms for Modern Parallel Computer Architectures*, Schultz, M.H., ed., Springer Verlag, pp 85-108, 1988.
- Ipsen, I.C.F. and Jessup, E.R., *Solving the Symmetric Tridiagonal Eigenvalue Problem on the Hypercube*, Research Report 548, Department of Computer Science, Yale University, July 1987.

  Accepted for publication in *SIAM Journal on Scientific and Statistical Computing*.
- Lam, J., *An Efficient Simulated Annealing Schedule* (Ph.D. Thesis), Research Report 8818, Department of Electrical Engineering, Yale University, September 1988.
- Lam, J. and Delosme, J.-M., *An Efficient Simulated Annealing Schedule: Derivation*, Research Report 8816, Department of Electrical Engineering, Yale University, September 1988.

  Submitted to *Algorithmica*.
- Lam, J. and Delosme, J.-M., *An Efficient Simulated Annealing Schedule: Implementation and Evaluation*, Research Report 8817, Department of Electrical Engineering, Yale University, September 1988.

  Submitted to *Algorithmica*.
- Lam, J. and Delosme, J.-M., *An Adaptive Annealing Schedule*, Research Report 8608, Department of Electrical Engineering, Yale University, September 1986.
- Lam, J. and Delosme, J.-M., *Performance of a New Annealing Schedule*, Proc. 25th Design Automation Conference (DAC), Anaheim, CA, pp 306-311, June 1988.
- Lam, J. and Delosme, J.-M., *Simulated Annealing: A Fast Heuristic for Some Generic Layout Problems*, Proc. Int. Conf. on Computer-Aided Design ICCAD-88, Santa Clara, CA, pp 510-513, November 1988.
- Lam, J., Delosme, J.-M. and Sechen, C., *A New Simulated Annealing Schedule for Row-Based Placement*, Proc. Int. Workshop on Placement and Routing, Research Triangle Park, NC, May 1988.

- Wong, Y. and Delosme, J.-M., *Transformation of Broadcasting into Pipelining*, Research Report 544, Department of Computer Science, Yale University, June 1987.
- Wong, Y. and Delosme, J.-M., *Optimization of Computation Time for Systolic Arrays*, Research Report 651, Department of Computer Science, Yale University, September 1988.

  Submitted to *Journal of VLSI Signal processing*.
- Wong, Y. and Delosme, J.-M., *Broadcast Removal in Systolic Algorithms*, Proc. Int. Conf. on Systolic Arrays, San Diego, CA, pp 403-412, May 1988.
- Wong, Y. and Delosme, J.-M., *Optimization of Computation Time for Systolic Arrays*, Proc. 26th Allerton Conf. on Communication, Control and Computing, September 1988.

D. Scientific Personal Supported

Jean-Marc Delosme
Ilse Ipsen
Ata Arjomand
Jimmy Lam
Liz Jessup
Theo Omtzigt
Yi-Wan Wong
Jin-Hong Ma

Advanced Degrees Earned by Scientific Personel:
Jimmy Lam: Ph.D. in Computer Science, December 1988